DOCUMENT RESUME

ED 046 992 TM 000 355

AUTHOR Goolsby, Thomas M., Jr.
TITLE Important Considerations in Achievement Test Selection for Diagnosis and Guidance.
NOTE 8p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Achievement Tests, Correlation, *Curriculum Evaluation, Diagnostic Tests, *Educational Diagnosis, Grade 7, *Guidance, Scores, Statistical Analysis, Student Evaluation, *Test Selection
IDENTIFIERS Metropolitan Achievement Tests, Stanford Achievement Tests

ABSTRACT

The relationship between total scores and interrelationships among certain similarly named subtests of the Metropolitan and Stanford Achievement Tests are considered. The correlation of total scores for the two batteries was .89, indicating that the batteries measure essentially the same things. If one were interested in selecting a battery to simply rank order students, either battery would be appropriate. Evidence of uniqueness of similarly named subtests in the two batteries was noted by the range of .49 to .69 correlations. Uniqueness of this magnitude is important in selecting a battery for measurement of outcome of instruction when one is interested in fitting the test to the curriculum. This is statistical evidence indicating that responsible school personnel need to analyze the curriculum by subject areas, consider future goals and innovations, and administer various subtests of various batteries to appropriate instructional personnel when selecting an achievement battery for diagnosis of pupil progress by areas and making decisions about the curriculum. (Author)

ED 046992

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

IMPORTANT CONSIDERATIONS IN ACHIEVEMENT TEST SELECTION FOR DIAGNOSIS AND GUIDANCE

Thomas M. Goolsby, Jr.
University of Georgia

Authors and publishers of extensive achievement test batteries have usually presented validity evidence of "content" and/or "curricular" types. This validity evidence has been quite appropriate since a criterion considered to be better than the battery and/or its subtests would be extremely difficult to identify or define.

Considerations of relationships between total scores and interrelationships among subtests for any two achievement batteries should be given particular attention by those responsible for achievement test selection. For extensive achievement batteries, correlations of total test scores on any two will approach .90 and many times exceed that value. It is possible for the zero order relationships of total scores on two batteries to approach unity (1.00) when one considers the very high reliabilities of batteries approximating .95. On the basis of the .90 magnitude of relationship for total scores, one might decide that one battery is just as appropriate to use with a given population as another. This is probably true if one is not particularly interested in determining subject area emphases for follow-up activities and curriculum guidance values of different achievement test batteries. Users of achievement test batteries are generally more interested in follow-up activities and curriculum guidance features of achievement tests than for a simple rank ordering or placement of students on the basis of a total score. The present study presents evidence of internal

TM 000 355

uniqueness of two achievement test batteries important to diagnosis and guidance.

The present study was designed to:

1. determine the degree to which like named components (science, mathematics and so forth) of two achievement test batteries measure the same things.
2. determine the correlation of total test scores for the two achievement test batteries.
3. determine something of the rational appropriateness of achievement test batteries for a given population.

Procedures

The two achievement test batteries considered were the Metropolitan Achievement Tests (MAT) and Stanford Achievement Tests (SAT) designed for measurement of outcomes of instruction at the same grade level. Selected subtests which are named exactly the same in both batteries and others which are similarly named were considered.

The population used was a large junior high school in a suburban area of greater metropolitan Miami. The MAT and SAT were administered to the seventh grade students in the fall. Intercorrelations among and between certain of the measures and between total scores for the two batteries were obtained. Means and standard deviations were computed for certain subtests of the two batteries.

Results and Discussion

Total scores on the MAT and SAT correlated at the .89 level. At this level of relationship, it can be said that the two batteries measure essentially the same things. This outcome was expected. Even

though these two batteries apparently measure the same things, the evidence that follows shows marked internal and practical differences important to achievement testing.

Table 1 presents relationships between certain subtests of MAT and

---

Insert Table 1 about here

SAT. The relationships corrected for attentuation between similarly named subtests in MAT with those in SAT.had a range of .49 to .69. The differently named subtests had a range of .39 to .57. These ranges are, in general, typical of interrelationships corrected for attenuation for subtests in a single achievement battery such as MAT or SAT. Those subtests in a single battery for which this range is not typical are subtests in a battery for special purposes such as vocabulary or special subject area coverage known to be highly related to other components but deemed important to include for special curricular emphases.

The relationships in Table 1 for similarly named subtests in the two batteries are probably surprising to many of those who frequently deal with practical measurement problems such as achievement battery selection to measure outcomes of instruction, pupil progress, and uniqueness of a given curriculum. The evidence presented here is supportive of the fact of uniqueness of curriculum areas as measured by two achievement batteries designed to measure outcomes at the same grade level. The evidence is supportive of the need for a variety of batteries constructed to fit differenct curriculums by subtests in both content and/or method.

Table 2 presents the means and standard deviations for certain of

---

Insert Table 2 about here

---

the MAT and SAT subtests. In general, the MAT is more appropriate for the population of this study. The SAT subtests presented are, in general, somewhat too difficult. The reader should not, however, depend too much on these values. The MAT was selected at an earlier date for the population of this study by analyzing the curriculum by areas, considering future goals and innovations, and administering subtests of various batteries to appropriate instructional personnel responsible for instruction in a given subject area.

Table 3 presents multiple correlations of certain MAT and SAT sub-

---

Insert Table 3 about here

---

tests. These multiple correlations are not substantially different from the zero order relationships presented in Table 1.

REFERENCES

Metropolitan Achievement Test. New York: Harcourt, Brace and World, 1959.

Stanford Achievement Test. New York: Harcourt, Brace and World, 1963.

TABLE 1

Relationships Between Certain Sub-Tests of MAT and SAT (N = 335)*

	1	2	3	4	5
6	53*** (58)**	46 (51)	46 (54)	40 (44)	48 (52)
7	45 (49)	52 (58)	52 (62)	43 (48)	50 (55)
8	39 (44)	48 (55)	52 (64)	53 (61)	58 (66)
9	39 (44)	44 (50)	47 (51)	50 (56)	59 (66)
10	49 (56)	55 (64)	57 (70)	49 (57)	60 (69)

1--SAT-Arithmetic Computation
2--SAT-Arithmetic Concepts
3--SAT-Arithmetic Application
4--SAT-Science
5--SAT-Social Studies
6--MAT-Arithmetic Computation
7--MAT-Arithmetic Problem Solving
8--MAT-Science
9--MAT-Social Studies Information
10--MAT-Social Studies Skills

*Decimals Omitted
**Zero order correlations corrected for Attenuation
***Zero order correlations

TABLE 2

Means and Standard Deviations for Certain MAT and SAT Subtests (N = 335)

	k*	r_{xx}**	Means	S. D.
SAT-Arithmetic Computation	41	.87	13.61	5.45
SAT-Arithmetic Concepts	40	.82	15.69	5.36
SAT-Arithmetic Application	36	.77	11.17	4.43
SAT-Social Studies	92	.89	40.47	14.01
SAT-Science	60	.88	25.06	8.88
MAT-Arithmetic Computation	45	.91	20.90	8.04
MAT-Arithmetic Problem Solving	48	.92	22.24	8.11
MAT-Science	55	.83	27.08	9.85
MAT-Social Studies Inform.	60	.89	27.77	9.91
MAT-S. S. Study Skills	40	.84	19.75	6.76

*Number of items
**Corrected split test reliability coefficient estimates reported by the authors of the tests.

TABLE 3

Multiple Correlations of Certain MAT and SAT Subtests (N = 335)*

7 vs 2 + 3 = 57

6 vs 2 + 3 = 51

2 vs 6 + 7 = 52

3 vs 6 + 7 = 52

5 vs 9 + 10 = 65

*Decimals Omitted

2---SAT-Arithmetic Concepts
3---SAT-Arithmetic Application
5---SAT-Social Studies
6---MAT-Arithmetic Computation
7---MAT-Arithmetic Problem Solving
9---MAT-Social Studies Information
10---MAT-Social Studies Skills